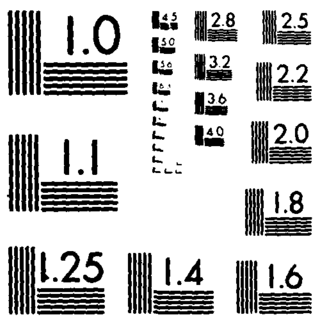
AD-A111 856 TELEDYNE GEOTECH ALEXANDRIA VA SEISMIC DATA ANALYSIS--ETC F/G 17/10
DISE, AN INTERACTIVE DISCRIMINATION PROGRAM FOR SEISMIC EVENTS.(U)
DEC 81 D VON SEGGERN F08606-80-C-0017
UNCLASSIFIED SDAC-TR-81-17 VSC-TR-82-4 NL

1 OF 1
AD
A111856

END
DATE
FILMED
4-82
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

(2)

VSC-TR-82-4

DISE, AN INTERACTIVE DISCRIMINATION PROGRAM FOR SEISMIC EVENTS

D. von Seggern

Seismic Data Analysis Center
Teledyne Geotech
314 Montgomery Street
Alexandria, Virginia 22314

18 DEC 1981

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Monitored By:
VELA Seismological Center
312 Montgomery Street
Alexandria, VA 22314

Sponsored by

The Defense Advanced Research Projects Agency (DARPA)

DARPA Order No. 2551

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER VSC-TR-82-4	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) DISE, AN INTERACTIVE DISCRIMINATION PROGRAM FOR SEISMIC EVENTS		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER SDAC-TR-81-17
7. AUTHOR(s) D. von Seggern		8. CONTRACT OR GRANT NUMBER(s) F08606-80-C-0017
9. PERFORMING ORGANIZATION NAME AND ADDRESS Teledyne Geotech 314 Montgomery Street Alexandria, Virginia 22314		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS VT/0707/B/PMP
11. CONTROLLING OFFICE NAME AND ADDRESS VELA Seismological Center 312 Montgomery Street Alexandria, Virginia 22314		12. REPORT DATE 12/18/81
		13. NUMBER OF PAGES 37
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Defense Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, Virginia 22209		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Seismic Event Discrimination

Interactive Discrimination

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

DISE (Discrimination and Identification of Seismic Events) is an interactive computer program with graphics support and currently runs on a VAX-11/780 computer at the SDAC. Using various commands which are available, the seismic analyst may employ location data or waveform measurements to identify unknown events. Groups of epicenters may be formed, and a lower level of subgroups is formed when particular stations or variables are selected for discrimination purposes. The program supports two basic approaches to event identification

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

using waveform-derived data: multivariate discriminant functions or multivariate clustering.

DISE, AN INTERACTIVE
DISCRIMINATION PROGRAM FOR SEISMIC EVENTS

SEISMIC DATA ANALYSIS CENTER REPORT NO.: SDAC-TR-81-17

AFTAC Project Authorization No.:	VELA VT/0707/B/PMP
Project Title:	Seismic Source Identification Research
ARPA Order No.:	2551
Name of Contractor:	TELEDYNE GEOTECH
Contract No.:	F08606-80-C-0017
Date of Contract:	1 May 1980
Amount of Contract:	\$144,164
Contract Expiration Date:	31 March 1981
Project Manager:	Robert R. Blandford (703) 836-3882

P. O. Box 334, Alexandria, Virginia 22313

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

A

ABSTRACT

DISE (Discrimination and Identification of Seismic Events) is an interactive computer program with graphics support and currently runs on a VAX-11/780 computer at the SDAC. Using various commands which are available, the seismic analyst may employ location data or waveform measurements to identify unknown events. Groups of epicenters may be formed, and a lower level of subgroups is formed when particular stations or variables are selected for discrimination purposes. The program supports two basic approaches to event identification using waveform-derived data: multivariate discriminant functions or multivariate clustering.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

The identification of the nature of seismic events in a seismological context is based on the examination of location coordinates and depth and on inferences from the recorded waveforms. These two types of data, i.e. "location" and "waveform", are distinct and are employed in different manners in the decision process. Although they complement and reinforce one another, each may, by itself, be sufficient to make a fairly reliable decision on the nature of an event. Ideally, they will be used together; and any complete analysis procedure will include both.

A rational strategy for identification of unknown seismic events is outlined in Figure 1. This strategy presumes that location information and certain waveform measurements are initially available to the analyst. The outcome of the analysis is a decision on the event type, if possible, and a confidence measure placed on that decision. The reader should not concentrate on particular procedural statements in Figure 1 because these will be made clear in the following section; rather the overall scheme should be examined. This scheme involves many branches and loops characteristic of complex decision processes. In fact many more possible paths than exhibited here would exist in an operational sense provided a sufficiently flexible program existed for implementing the concepts of Figure 1.

This report describes an interactive program which implements the scheme of Figure 1 and which has been coded and installed on a VAX 11/780 computer at the Seismic Data Analysis Center. This program has the title "Discrimination and Identification of Seismic Events", or briefly "DISE", an acronym which emphasizes the role of probability and statistics in the decision process. This program fulfills two important current requirements:

1) provision of an interactive and graphical program for studying and optimizing the event classification properties or features of seismic data and

MODULE 1

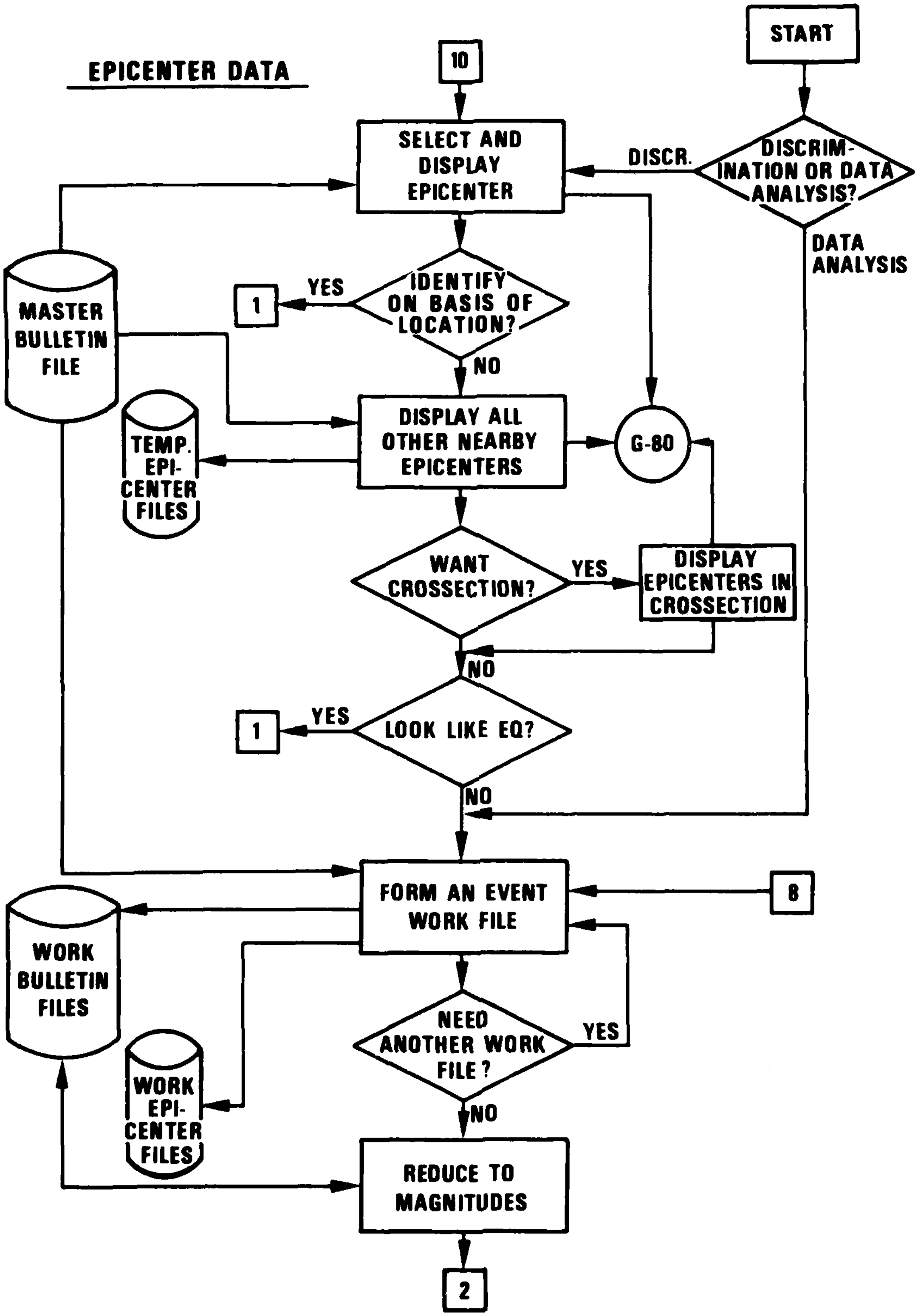


Figure 1. [illegible] ow in an interactive discrimination mode.

MODULE 2

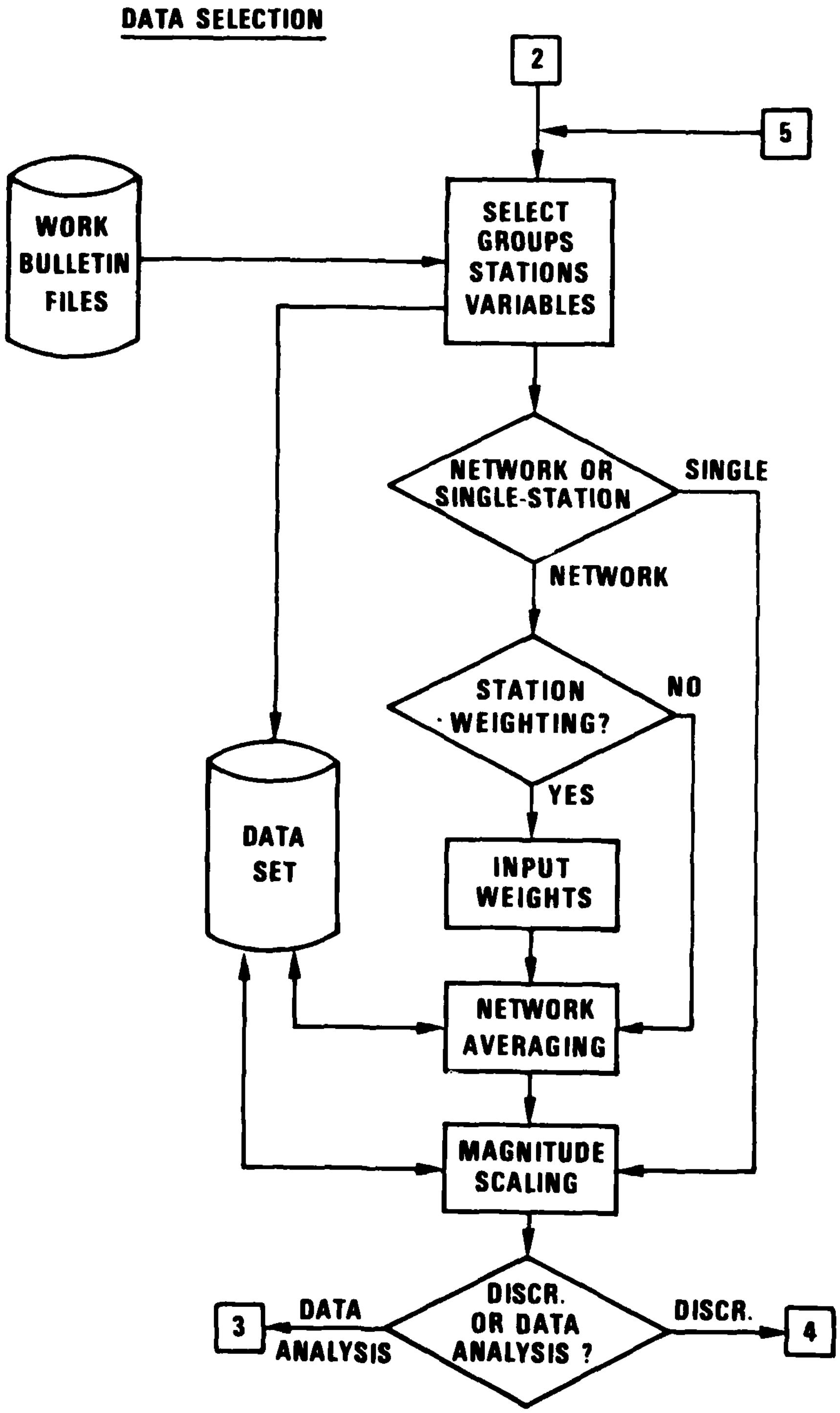


Figure 1. continued. Logic flow in an interactive discrimination mode.

MODULE 3

DATA ANALYSIS

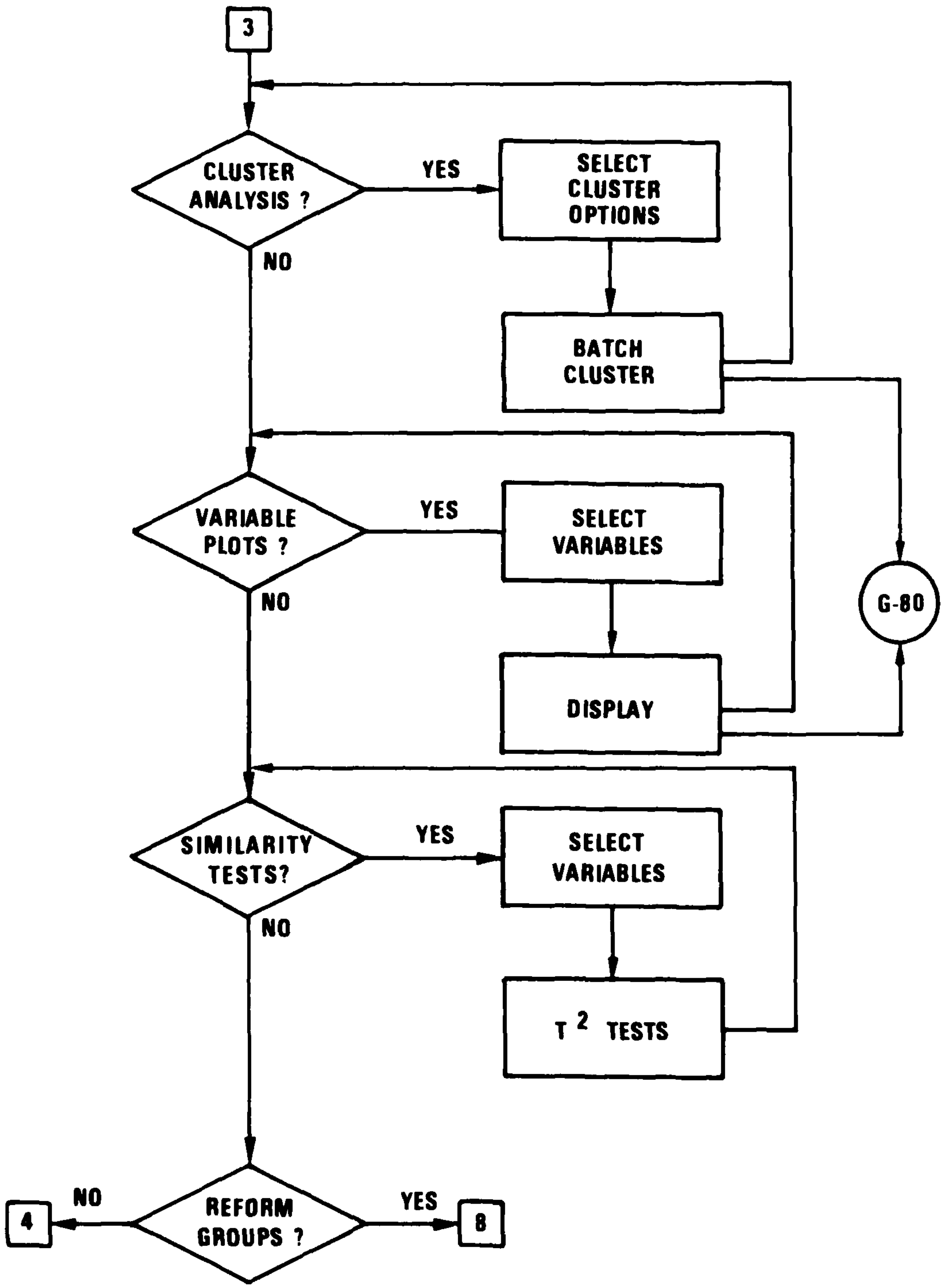


Figure 1. continued. Logic flow in an interactive discrimination mode.

MODULE 4

DISCRIMINATION

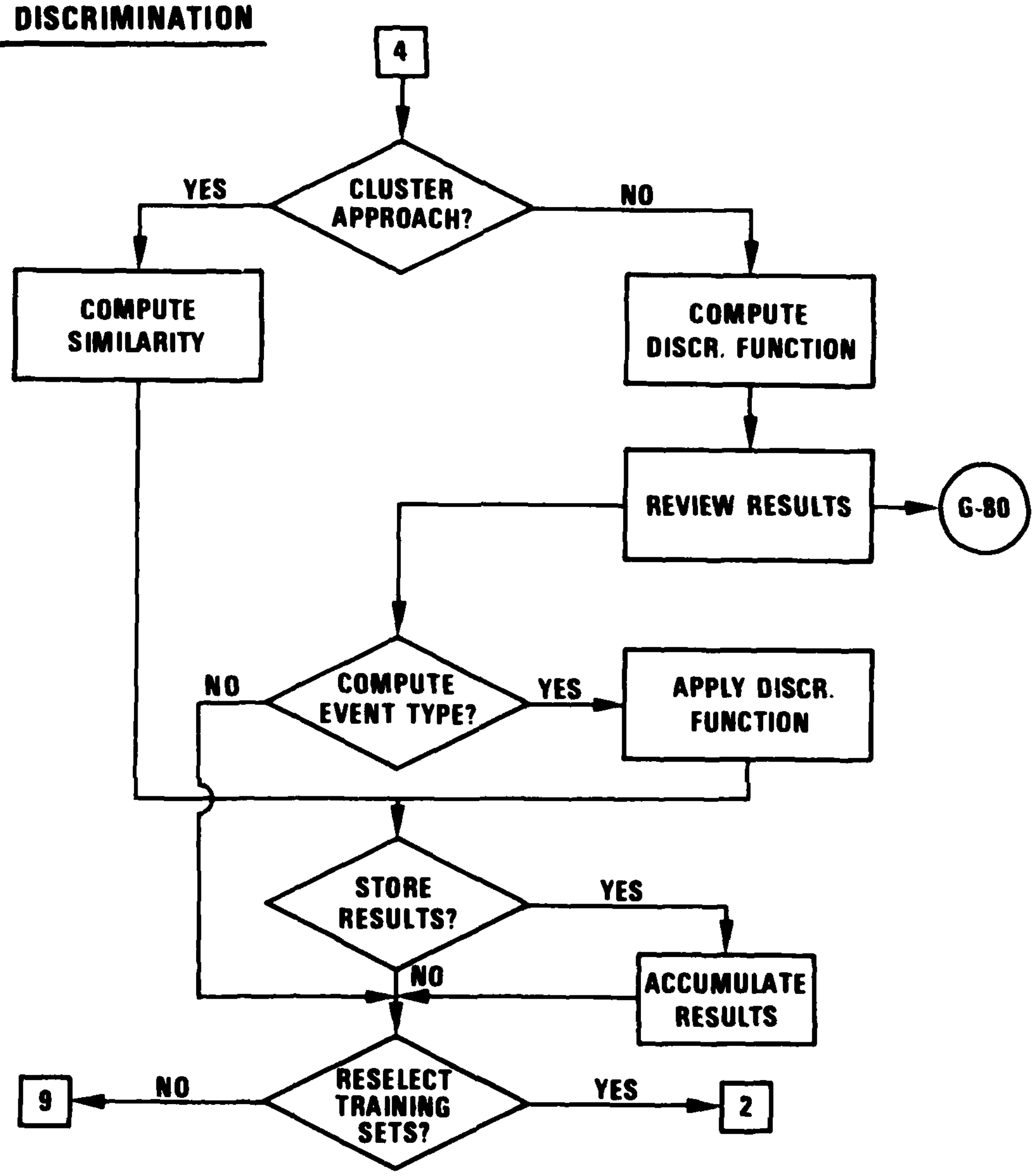


Figure 1. continued. Logic flow in an interactive discrimination mode.

MODULE 5

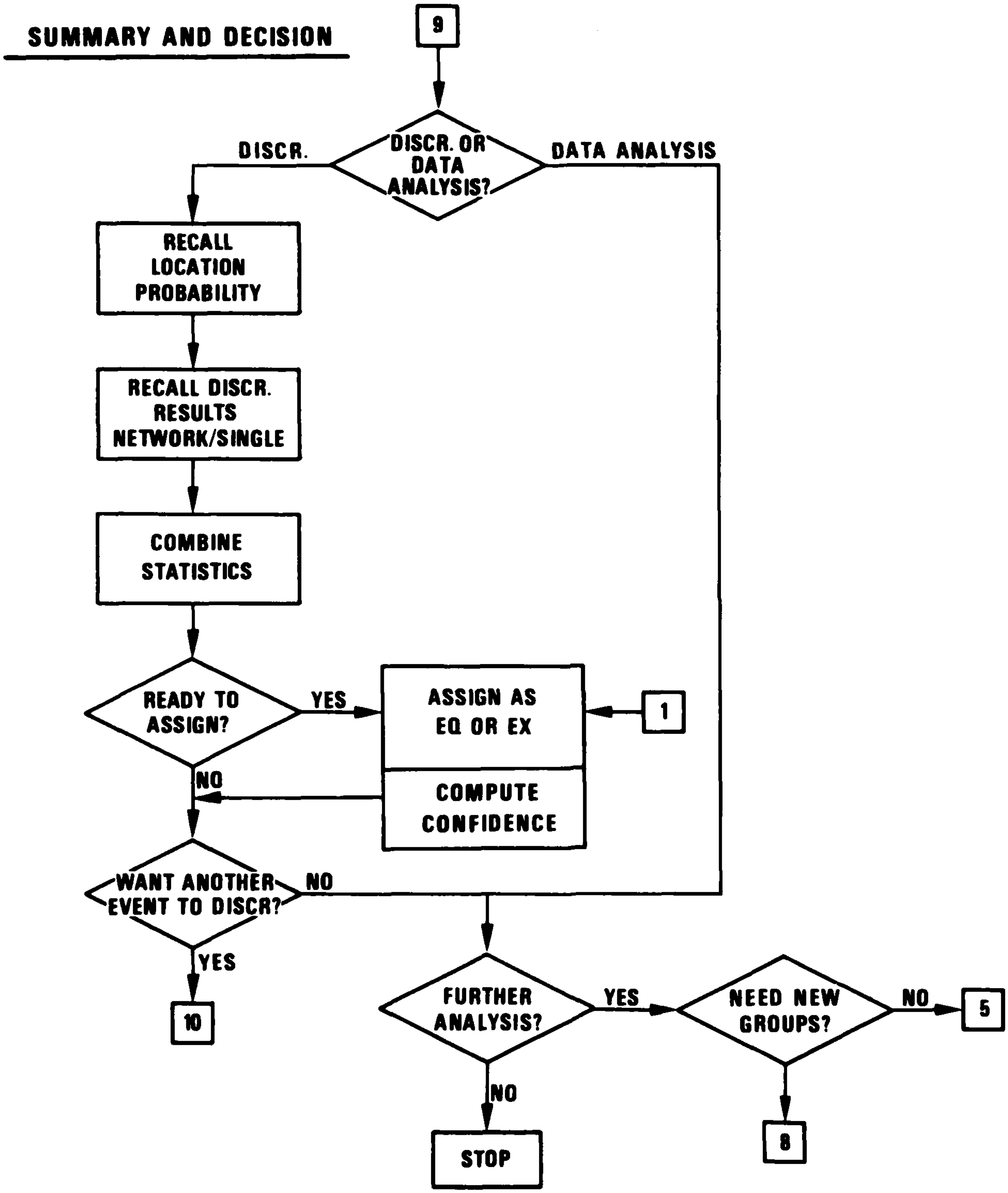


Figure 1. continued. Logic flow in an interactive discrimination mode.

2) provision of a prototype program to routinely identify events contained in a seismic bulletin in an efficient manner using all appropriate seismic data input.

The background for Figure 1 consists of twenty years of research into seismic event identification under the VELA-Uniform program, as summarized in Rivers et al. (1981). That research is composed largely of disjointed studies on limited data bases using one particular discrimination technique; but more recently involved multivariate discrimination, culminating in a ponderous experiment on 133 Asian events by several contractors. A thorough evaluation of this recent experiment is also contained in Rivers et al. (1981). The scheme of Figure 1 embodies the statistical techniques used in the Asian event discrimination experiment and in fact would allow for a duplication of many of the results of that experiment, given the same data base.

Note that the procedure outlined in Figure 1 does not include the actual location of a seismic event or waveform measurements. These functions occur outside the formal framework of the present DISE program; however, future versions of the program will be linked into a larger seismic analysis program currently being coded for the VAX under separate contract, and the capability of refining locations or remeasuring waveform variables within the discrimination procedure would then exist. This full capability is deemed essential to a truly efficient and reliable discrimination effort.

The remainder of this report consists mainly of a description of the DISE program and its environment. The description of the program is presented in a format which lists and explains the purpose of each command available to an analyst using the program. It is intended to be a functional description of the program as it existed at the time this report was prepared. Detailed documentation of the actual coded modules which comprise the DISE program is available from the SDAC on request.

DESCRIPTION OF THE DISE PROGRAM

Functional Summary

The DISE computer program is a seismic analysis tool to:

1) study and optimize the classification properties or features of seismic data from natural earthquakes and artificial seismic events, i.e., explosions,

2) identify events contained in a seismic bulletin as earthquakes or explosions, or otherwise, in a rapid and routine manner.

The procedure for accomplishing these tasks is embedded in an interactive analysis mode with graphics aids and statistical support routines. The basic input to the program is a seismic bulletin listing event hypocenters and their associated seismic arrivals at various recording stations. Measurements on these arrivals are assumed to be present in this bulletin; and the program, as presently configured, does not allow for making further measurements or changing those already available.

The DISE program permits the definition of event groups within the original set of events. The defining criteria are:

a) location,
b) depth,
c) magnitude,
d) event type,
e) date.

Editing procedures are available for refining the event list corresponding to a given group.

Furthermore, subgroups can be defined from their respective groups by the following criteria:

a) station,
b) variable.

The DISE program permits application of the following discrimination techniques:

a) event location,
b) event depth,
c) multivariate discrimination using waveform-derived data,
d) clustering using waveform-derived data.

The DISE program provides the following interactive graphics displays to aid the discrimination process:

1) epicenter and station locations on coastline and political boundary map background,

2) hypocenters projected on vertical cross-sections,

3) scattergrams and histograms for discrimination variables.

The interfacing of the functioning software and hardware components of DISE is shown in Figure 2. In addition to the VAX itself, the other hardware components which are employed are the Graphicus-80 CRT provided by Vector Automation, Inc., an Intercolor A/N terminal, a Talos data tablet, a DEC printer, and a Versatec hardcopy unit. The analyst controls the flow of data and processing through the A/N terminal by a set of commands with qualifiers. He also controls the data displays on the Graphicus-80 through its keyboard and the data tablet.

Data Organization

Knowledge of how the DISE program organizes and stores the data is essential to a proper understanding of the commands available to the analyst. The data base at the beginning consists of measurements made at certain stations for certain events. Thus, in statistical terms, there are multiple variables (seismic measurements) for multiple cases (events). The analyst first designates groups of cases (events) on the basis of their location, magnitude, or event type (e.g. explosion and earthquake). Events of unknown type will be formed into another group. Then for each group, the analyst may want to define various subsets of variables. These variables may be chosen from only one station or from several stations, resulting in an averaging procedure over the stations

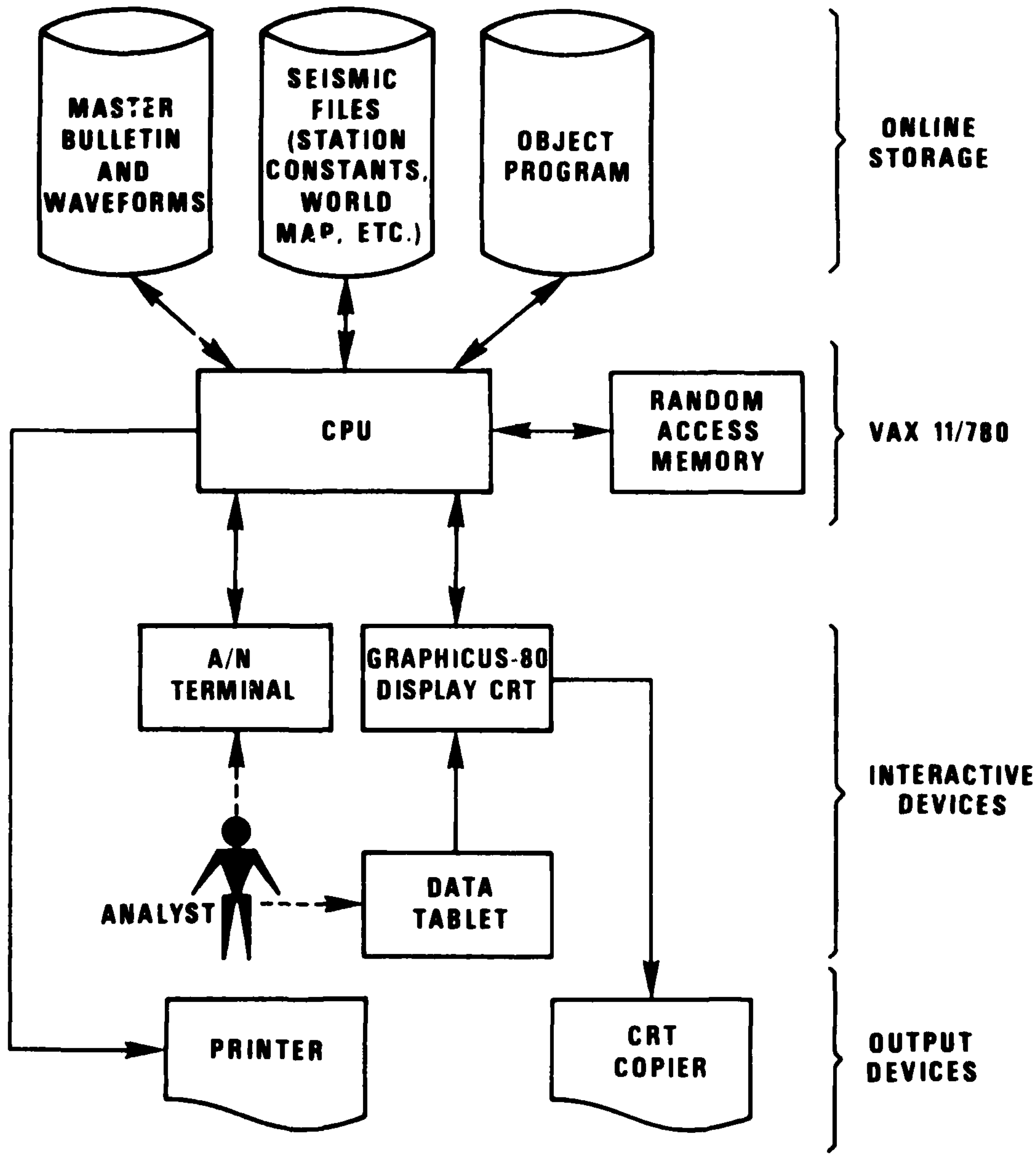


Figure 2. Hardware and software components pertaining to the DISE program.

to produce network variables. Each definition of a subset of variables involves both variable and station selection and creates a subgrouping of data within the original data available for the group. The subgroup contains all the cases (events) of the original group but only a clearly defined subset of the variables. Multiple subgroups can be created for each group depending on the analysis strategy.

While location analysis will be handled at the group level, statistical discrimination procedures will be performed at the subgroup level.

Interactive Discrimination Commands

The program flow is controlled by analyst commands entered through the A/N keyboard terminal. Presently the command syntax is unflexible, and command strings consist of variables separated by commas. A list of the available commands and a brief description of their intended usage is given in Table I. A detailed description of each command in Table I is contained in the following subsections.

TABLE I

Interactive Discrimination Commands

Command Name	Purpose
INIT	Initialize the program and set up data base.
GETE	Select an event to discriminate from master event set.
SELE	Get all epicenters from master event set within prescribed selection criteria.
MAPE	Display epicenters on map background.
CROS	Perform a transformation to put epicenters on a vertical cross section.
FRMG	Create a working set (group) of events taken together as a group.
TRSG	Transfer a working set (group) into temporary set.
DELG	Delete a working set (group).
SUMG	Summarize in tabular format the data availability for each group.
SDAT	Select data according to event groups, stations, and variables for use in statistical analysis.
SCAL	Apply scaling relations to remove magnitude-related bias of data among groups.
SIML	Compute similarity measure of a given event to previously formed groups, or similarity of groups.
CLUS	Perform cluster analysis on selected groups of events.
SPLT	Reform the event groups on the basis of clustering results.
XYPL	Form and display plots of any two selected variables.
HGPL	Display histogram of a given variable from a selected subgroup.
CDIS	Compute the multivariate discriminant function between two subgroups of data.
ADIS	Apply discriminant function to unknown event(s).
SDIS	Sum discriminant functions over stations.
LSTG	Display the attributes of all current groups.
LSTS	Display the attributes of all current subgroups.
LSTD	Display information pertaining to current discrimination functions.

INIT: Program Initialization

Presently, this command is automatically invoked at the start of the program execution because the data sets are prepared external to the program. One of these data sets is the master variable file which is organized as shown in Table II and retained on disk. All data for a single event is grouped sequentially in this file. Another data set is read into core from disk and is organized as shown in Table III into a common area. Note that a pointer to the master variable file, which is random-access, is defined for each event.

TABLE II

FORMAT OF THE MASTER VARIABLE FILE

Parameter	Storage Mode
Event Number	I*2
Station Number	I*2
Variable Number	I*2
Amplitude	R*4
Magnitude	R*4
Distance to source	R*4
Azimuth from source	R*4

TABLE III

FORMAT OF THE MASTER EVENT COMMON AREA

Parameter	Storage Mode
Event Number	I*2
Pointer to master bulletin	I*2
Pointer to master variable file	I*2
Flag for temporary display	I*2
Group Number	I*2
Cluster Number	I*2
Event type	L*1
Date	I*4
Origin Time	I*4
Latitude	R*4
Longitude	R*4
Depth	R*4
Magnitude	R*4
Major semiaxis length	R*4
Minor semiaxis length	R*4
Aximuth of major axis	R*4

GETE: Select one epicenter

This function selects a single hypocenter for the purpose of identification and displays all the available data for it to the analyst. Input consists of analyst-supplied date and origin time. A search is made through the master event set for the event, it is flagged as a "temporary event set" and the epicenter data for the event is displayed at the analyst's terminal.

SELE: Select a group of epicenters

This function enables the analyst to select a group of events on the basis of their hypocenter attributes. Such a group will often, but not always, assume the role of a "training" set for the purpose of identifying unknown events. The analyst supplies latitude, longitude, depth, magnitude, and date limits to be used as selection criteria. The analyst also specifies event type, either explosion or earthquake, or otherwise, with the default being unidentified events. A search is made through the master event set for those events satisfying the selection criteria, and selected events are flagged as a "temporary event set". A summary of the selection results is listed on the A/N terminal; specifically, the analyst sees the number of events found and the actual events which satisfied the specified selection criteria.

MAPE: Map of epicenters

This function presents the analyst with a graphics display of the epicenter(s) in the temporary event subset. They are displayed in plan view or on a map with background of coastlines and political boundaries. The analyst enters a map projection option, whether or not confidence limits on latitude and longitude are to be shown about the epicenters, and whether or not presently displayed epicenters are to be retained on the screen in addition to those of the new temporary event subset. The required spatial limits of the map are computed and the appropriate sections of the coastline and political boundary files are accessed to construct the map. A transformation on latitude and longitude, if required, is made to satisfy the analyst's projection option. The epicenter coordinates are similarly transformed before displaying. Confidence ellipses, if required, are read from the master event common area and displayed about the epicenter(s). The visual output consists of the graphics display. A map bookkeeping record is updated to reflect current map limits, map projection type, and subsets of data being displayed. A "pick" attribute is associated to each epicenter on the screen for later editing purposes.

CROS: Cross-section of epicenters

This function converts a plane view of hypocenters to a display on a plane normal to the earth's surface to provide visual aid in depth analysis. The analyst supplies the orientation of the plane to which the hypocenters are projected and the maximum depth of hypocenters to be displayed.

Let α be the specified angle of the projection plane from north ($0 < \alpha < \pi/2$). Let (θ, ϕ) be the colatitude and east longitude of the centroid of epicenters to be projected. Consider then the spherical triangle in Figure 3 with angles A, B, and C and sides a, b, and c (in radians). Let the projection of an epicenter at (θ_e, ϕ_e) be normal to the plane so that $C = \pi/2$. Using the law of sines:

$$\frac{\sin a}{\sin A} = \frac{\sin c}{\sin C}$$

Since C is a right angle, we have from spherical trigonometry:

$$\sin a = \sin c \cdot \sin A$$

or

$$a = \arcsin (\sin c \cdot \sin A)$$

Thus $\underline{a}$ is represented in terms of known quantities since $\underline{c}$ and $A = |\beta - \alpha|$ can be computed from (θ, ϕ) and (θ_e, ϕ_e). Again, from spherical trigonometry:

$$\sin b = \sin B \cdot \sin c$$
$$\cos B = \tan a \cdot \cot c$$

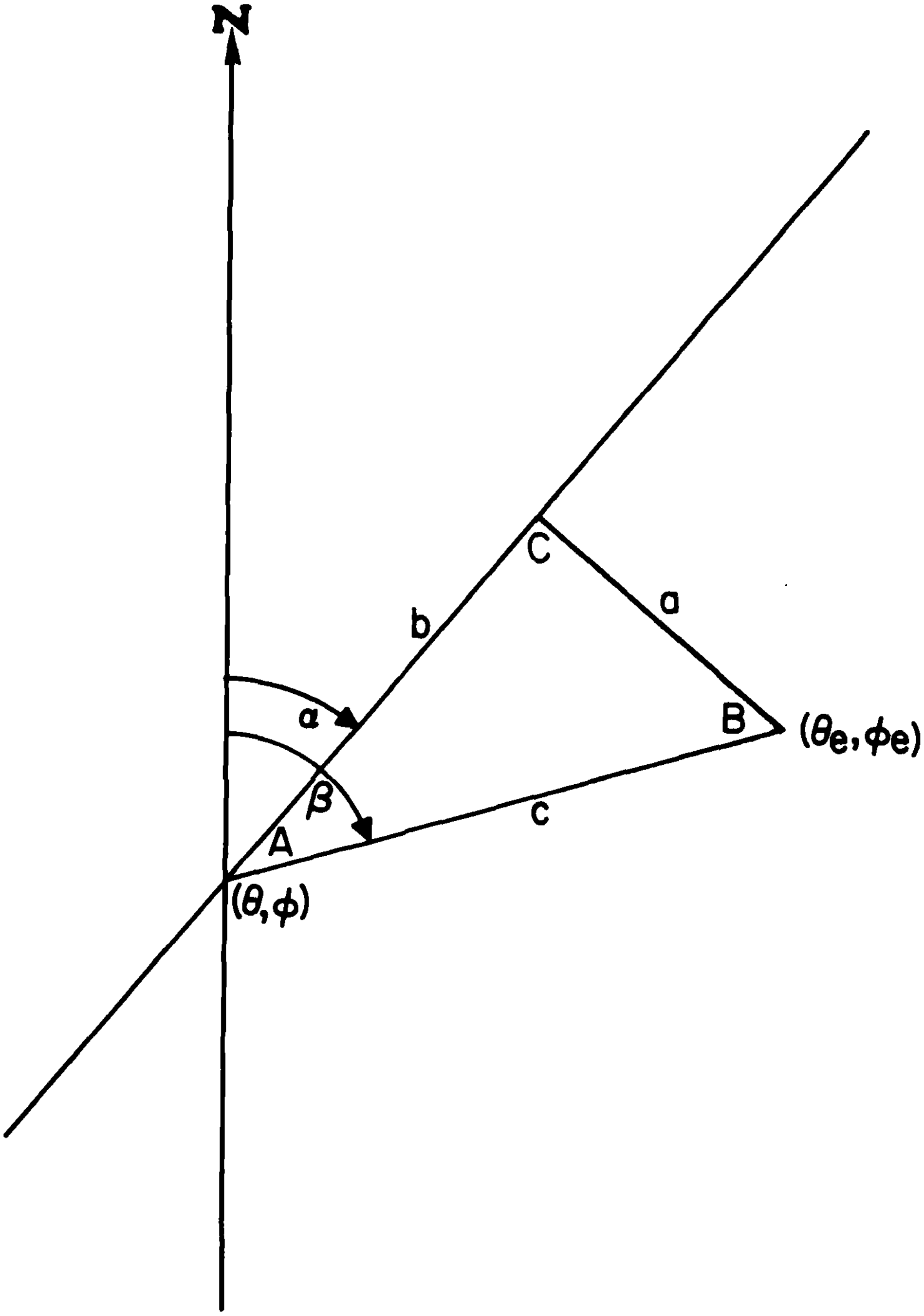


Figure 3. Method of projecting epicenters to a great-circle path.

Using $\cos^2 B + \sin^2 B = 1$ gives

$$\sin b = \sqrt{1 - \tan^2 a \cot^2 c \sin c}$$

or

$$b = \arcsin\left[\sqrt{1 - \tan^2 a \cot^2 c \sin c}\right]$$

This gives the horizontal distance along the plane where the given epicenter lies relative to the reference centroid point. With all epicenters projected in this manner, the cross-section view is displayed on the graphics screen. The confidence limits on depth and on the horizontal coordinate are displayed if requested. Since the projection plane lies at an orientation α from north and the confidence ellipse at an orientation γ, the appropriate confidence limit for the projected epicenter is the intersection of a line parallel to the plane with the ellipse as shown in Figure 4. Let $\delta = |\gamma - \alpha|$) and p and q be the major and minor semi-axis lengths of the ellipse, respectively. Then L, the projected confidence limit, is contained in the equation of the ellipse thus:

$$\left(\frac{L \cos \delta}{p}\right)^2 + \left(\frac{L \sin \delta}{q}\right)^2 = 1$$

Solving for L gives

$$L = \left(\frac{\cos^2 \delta}{p^2} + \frac{\sin^2 \delta}{q^2}\right)^{-\frac{1}{2}}$$

FRMG: Form group of events

This function formally designates a temporary event subset as a "group". Groups are logically associated events which will serve as training sets for later discrimination or whose features will be later compared analytically with those of other groups. A group number and symbol are provided by the analyst (the symbol serves to identify

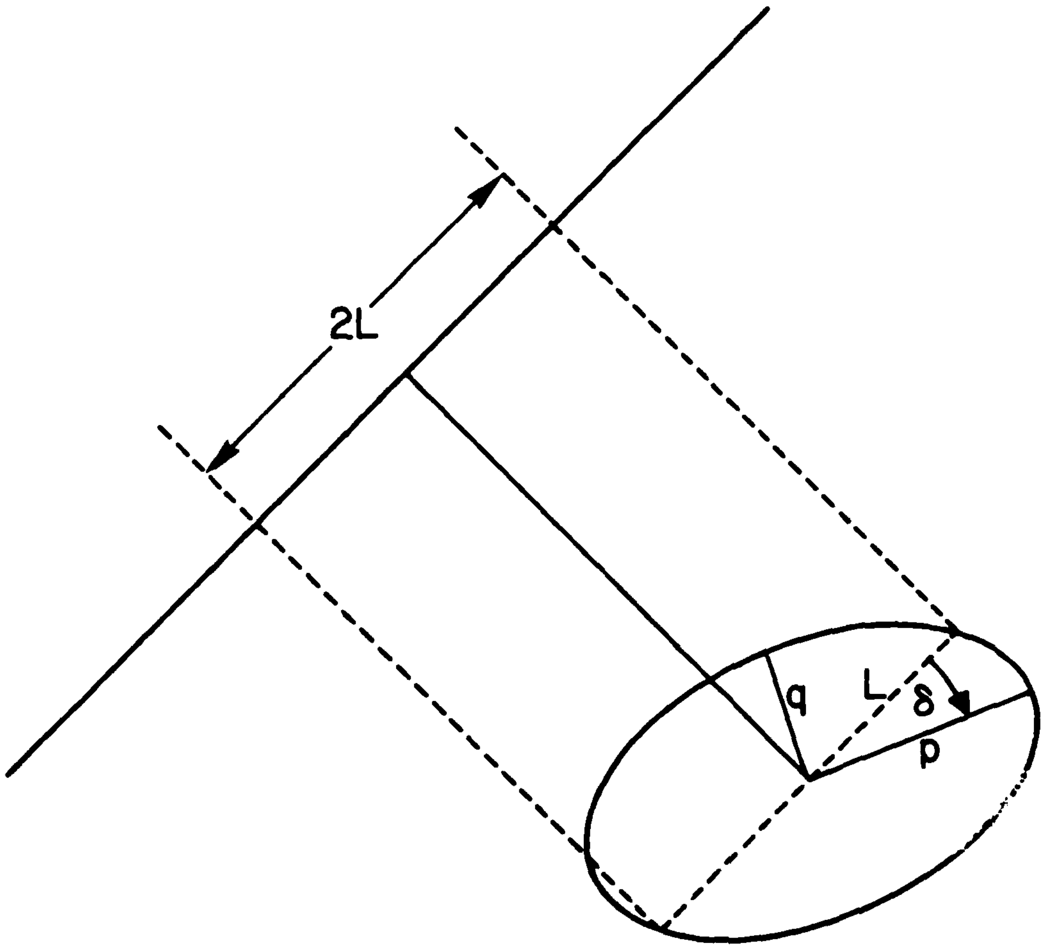


Figure 4. Method of projecting a confidence ellipse to a great-circle path.

members of this group throughout all later stages involving graphics displays). The analyst also enters an alphanumeric string to describe this group. Checking is required to prevent overwriting a previously-formed group, and the analyst is prompted to choose another group number if it already is used or to delete the previous group if desired. The group association number is set in the master event common area and the bookkeeping records on groups are updated to reflect the new group. Epicenter symbols on the map (if displayed) are changed from the temporary symbol to the designated group symbol.

TRSG: Transfer a group

This function transfers a chosen group back to temporary status for editing and display of epicenters. The temporary flags in the master event common area are set for those events belonging to the specified group. The analyst can subsequently reform the group with the command function FRMG or leave it unchanged.

DELG: Delete a group

This function deletes a specified group. All variables set by FRMG for this group are erased and any subgroups created through the SDAT command, as described later, are erased. Epicenters for this group, if displayed, are deleted from the graphics display.

SUMG: Summarize group data

This function summarizes the available discrimination data for each group with respect to station or variable. This enables the analyst to see exactly what data is available.

The available data for all specified groups is tabulated in a triply-indexed array (group, station, and variable). A summary table is prepared by counting flags for the variable(s) or station(s) requested by the analayst. Figure 5 illustrates the format for this table. In the first table (by station), the number of events in each group with the variable identified as "MB" is listed for all stations in the data base. In the second table (by variable), the number of events for which the station "ALE" appears with all of the possible variables in the data

STATION	VARIABLE = m_b GROUP						
	1	2	3	4	5	6	7
ALE	1	5	3	2	2	6	7
BLA	2	7	3	3	5	5	4
COM	2	8	4	4	4	3	2
.							
.							
.							

VARIABLE	STATION = ALE GROUP						
	1	2	3	4	5	6	7
M_s	5	5	6	3	2	2	3
m_b	6	2	1	7	6	5	2
FC	2	3	3	4	8	4	3
.							
.							
.							

Figure 5. Format of tables which summarize group data by station or by variable.

base is listed for each group.

SDAT: Formation of subgroups

This function creates subgroups from any existing groups. The analyst specifies the groups for which the subgroups will be made. He also enters stations and variables to be considered in the formation of the subgroups. The subgroup data sets contain variables characterizing the events in the groups, implying an averaging of magnitudes over stations to produce a network value if more than one station is requested. The subgroup organization of variables is appropriate for later input to various statistical or analysis modules. Subgroups pertaining to each of several stations can be formed by specifying only one station in successive subgroup formations.

Subgroup formation proceeds one group at a time. Within each group, the data for each event is searched for the specified variables and stations. If more than one station is specified, network averaging of variables is accomplished by the method of Ringdal (1976). Any event for which no actual measurement of a given variable is available among the specified stations will have an "estimated" value placed on this given variable if noise measurements are available. This value is at the 5% probability level (thus an upper bound) of the magnitude for non-detection of the associated seismic phases at all stations given their noise levels (again using formulas in Ringdal). A flag is set for the variable to indicate that no actual signal measurements were involved in computing it. If no data whatsoever is available, the flag is set to another value and a zero assigned to the variable. The subgroup data is retained in core and pointers created to identify the starting locations for each subgroup. The arrangement of the subgroup data is illustrated in Figure 6. Entries will be made in a subgroup information table to indicate the parent group, the number of events in the subgroup, and the stations and variables represented in the subgroup.

SCAL: Variable scaling

This function removes bias present in the variables due to event

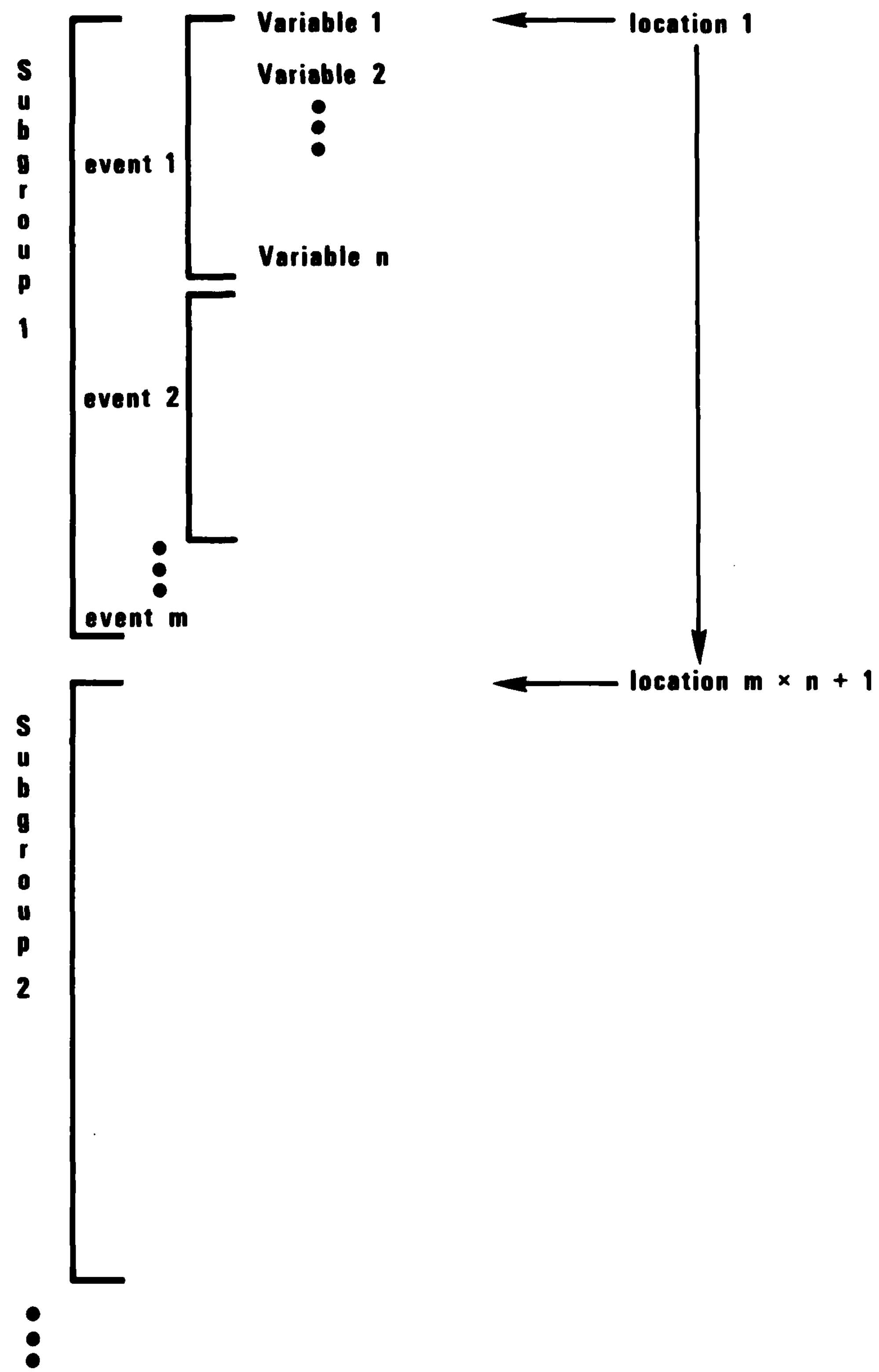


Figure 6. Storage arrangement for subgroup data array.

size. This is required when, for instance, the training sets of explosions and earthquake have significantly different sample means of bodywave magnitude. Empirical scaling relations are invoked to remove the bias. The process may also be termed "normalization" of the variables.

The analyst specifies the subgroup numbers for which the scaling is to be performed. A set of scaling coefficients is available in a common area. The scaling will be done in reference to a particular variable, body-wave magnitude being the recommended reference. Scaling formulas are of the linear type:

$$x' = ax + b$$

where x is the value of the variable to be scaled and a and b are coefficients. Some variables, by their nature, require no scaling (e.g., complexity) and will be unchanged in this processing. The output is the replacement of the variables in the subgroup data array by the scaled ones. A flag will be set in the subgroup information table to indicate that scaling was performed.

SIML: Similarity of groups

This function provides the analyst with a rapid quantitative measure of the similarity between two groups of events. Optionally, one group may contain only one event; and in this case the likelihood of the event belonging to the specified group is computed. This function is a simple, but less informative, means of classifying events than the application of full multivariate discrimination, to be described later.

In the case of more than one event in both groups, the Mahalanobis distance

$$r^2 = (m_1 - m_2)^T \Sigma^{-1} (m_1 - m_2)$$

is computed. Here $\bar{m}_1$ and $\bar{m}_2$ are the vectors of group means of the variable vectors x, computed thus

$$m_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} x_{1i}$$

$$m_2 = \frac{1}{N_2} \sum_{i=1}^{N_2} x_{2i}$$

where N_1 and N_2 are the number of events in each group. Σ is a pooled covariance matrix computed thus

$$\Sigma = \frac{(N_1 - 1)\ \Sigma_1 + (N_2 - 1)\ \Sigma_2}{N_1 + N_2 - 2}$$

from the two group covariance matrics Σ_1 and Σ_2.

The quantity

$$T^2 = \frac{N_1\ N_2}{N_1 + N_2}\ r^2$$

called the Hotelling's T^2 statistic, can be converted to an F statistic thus (Morrison, 1976, p. 137)

$$F = \frac{N_1 + N_2 - p - 1}{(N_1 + N_2 - 2)p}\quad T^2$$

where p is the number of variables. This quantity is compared against standard cumulative F tables with p and $N_1 + N_2 - p - 1$ degrees of freedom, yielding a probability P. The probability of the two groups belonging to the same population is then given as 1-P.

In the case of only one event in one of the groups, the test statistic becomes simply

$$\chi^2 = (x_1 - m_2)^T\ \Sigma_2^{-1}(x_1 - m_2)$$

where Σ is based on the second group only. This quantity is compared to standard χ^2 tables with p degrees of freedom, yielding a probability P. The probability of the single event belonging to the second group is then 1-P.

CLUS: Clustering of events

This function clusters seismic events based on specified measurement variables. The computational process is contained in the IMSL subroutine OCLINK (IMSL,1980); and it is beyond the scope of this report to describe it fully. Clustering is an alternative approach to the linear discrimination function as a means of classifying events. It is also likely to be invoked to determine the homogeneity of training event sets or the source-region effects on discrimination variables.

The analyst inputs the final number of clusters to be formed from all the events and the subgroup numbers from which these events are to be taken. Options for different clustering algorithms are also input. The results of the clustering are presented on the A/N terminal. A number is set in the master event common area to indicate the cluster group to which each event was amalgamated.

SPLT: Reformation of groups

This function forms new event groups based on the clustering results. Essentially it accomplishes what several passes through the function FRMG described above would do.

XYPL: Scattergrams

This function provides a scattergram on the graphics terminal of any two specified variables as an analysis aid in determining group characteristics. Events with outlying variable values can be readily identified in this manner. Variables from more than one group can be displayed simultaneously, and the symbols will be those originally assigned to the groups by the FRMG command described above.

HGPL: Histograms

This function provides a histogram on the graphics terminal for any specified variable. This allows the analyst to see the distribution of that variable for a given group, or optionally, for two groups such as explosions and earthquakes.

CDIS: Multivariate Discrimination

This function forms the multivariate discriminant function between two or more groups ("training sets") of seismic events. The processing consists largely of the IMSL subroutine ODNORM (IMSL, 1980). All the variables in the specified subgroups will be used. Those events having an incomplete set of variables (the case of flags indicating no data available) will be deleted from the training set prior to computation. The important computed quantities are the group classification vectors

$$D_k = \Sigma^{-1} x_k \qquad k = 1, \text{\# groups}$$

and their associated constants

$$C_k = -\tfrac{1}{2} x_k^T \Sigma^{-1} x_k \qquad k = 1, \text{\# groups}$$

where x_k and Σ are as described for SIML above. The D_k vectors and C_k values are stored in a discrimination table, which indicates the subgroups involved. An indication of the power of each variable for discrimination among groups is computed through IMSL subroutine ODFISH (IMSL, 1980) and displayed on the A/N terminal.

ADIS: Apply Discrimination

This function applies the group classification functions computed as described above to one or more events of unknown type. The probability of belonging to each group is computed. The classification scheme is illustrated in Figure 7 for the case where network variables are used; i.e., the discrimination variables are averaged over the network of stations when the subgroup data is formed in SDAT.

The analyst has the option to specify non-equal prior probabilities for the different groups to which the unknown event(s) is being classified.

The a posteriori probabilities

$$P_k(x) = \frac{\exp(x \cdot D_k + C_k)}{\sum_{i=1}^{M} [\exp(x \cdot D_i + C_i)]}$$

are computed for classifying the unknown event with variable vector x to each of M groups. These probabilities are displayed to the analyst. The classification values for each event, as represented by

$$x \cdot D_k + C_k \qquad k = 1, M \text{ groups}$$

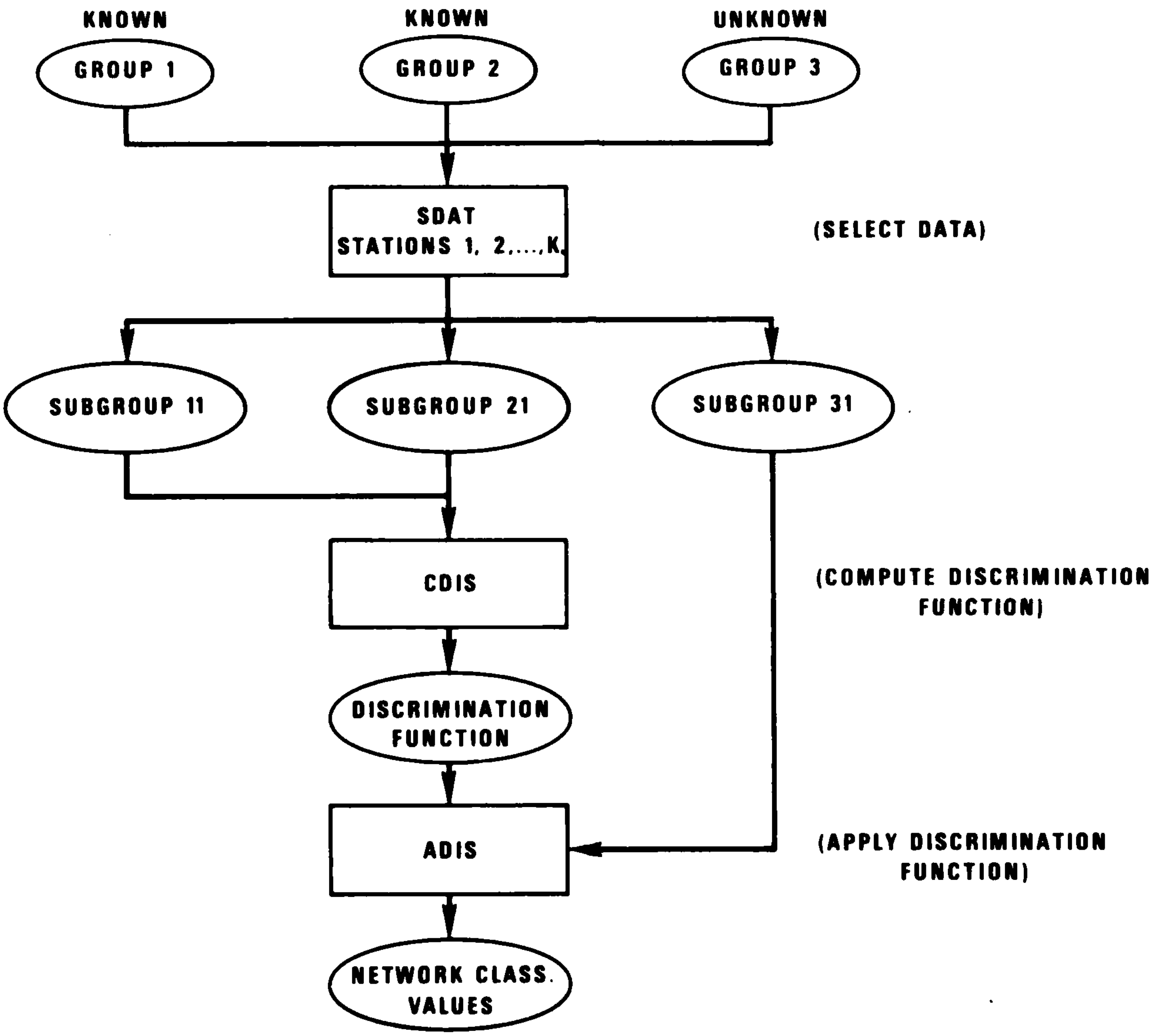


Figure 7. Scheme for event classification using network averaged variables.

will be stored in a classification table for later reference during execution of the SDIS command described next.

SDIS: Sum discriminants

Sometimes the analyst may wish to compute classification functions for individual stations of a network and then apply these to the unknown events separately. The network classification values are then the sum of the individual station values. This scheme is illustrated in Figure 8, and this command performs the final step of that illustration. Figure 8 should be compared to Figure 7, and a discussion of the relative merits of the two discrimination approaches is given in Rivers et al. (1981).

The analyst can optionally specify weights w_j for each of the L stations used. The summed classification values are given by

$$\sum_{j=1}^{L} w_j[x_j \cdot D_{jk} + C_{jk}]$$

For each unknown event, a posteriori probabilities are computed as described for the ADIS command, except the above expression replaces $x \cdot D_k + C_k$.

LSTG: List Groups

This command produces a listing on the A/N terminal of the attributes of each group formed in the analysis session; i.e., the selection criteria, the event type, and the number in each group.

LSTS: List Subgroups

Similar to LSTG, this command produces a listing of all the subgroups formed in the analysis session, with the stations and variables which defined them.

LSTD: List Discriminants

This command produces a listing on the A/N terminal of all the pertinent information concerning computed classification functions.

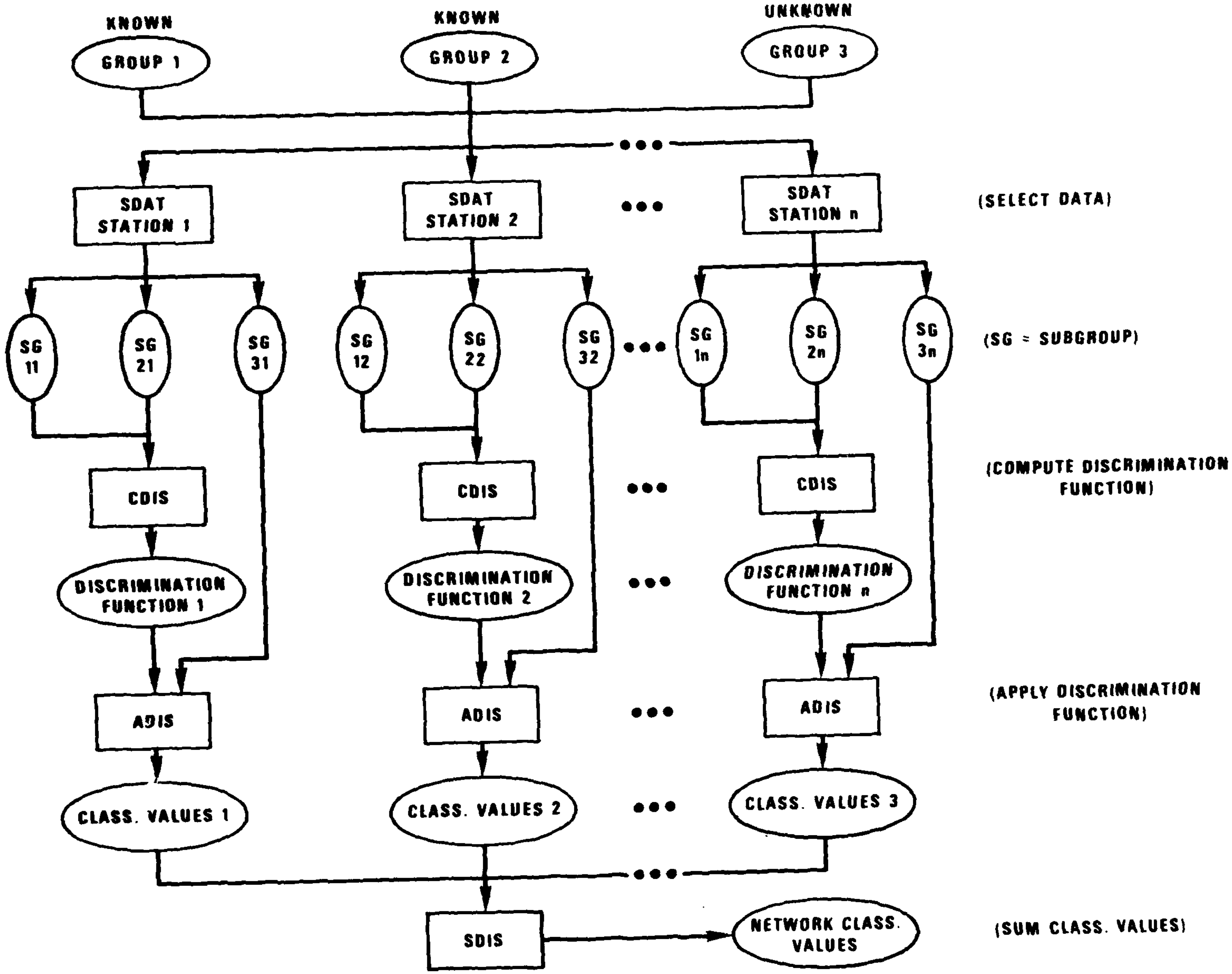


Figure 8. Scheme for event classification using network averaged classification values.

DATA BASE AND SUPPORT

Master Bulletin File

The Regional Event Location System (RELS) being designed under separate contract at the SDAC will produce an output, called the EAF (Event Arrival File), which can properly be termed a seismic buletin. It will contain event information and data for seismic phases associated to each event. The format of the EAF is as presented in "Data Base Specifications--Seismic Research Information System, " available at the SDAC. It is not yet known how the RELS bulletin will be mass stored. In any case, the analyst will only want a portion of it for any given session. It is foreseen then that an offline job will prepare a subset of the full bulletin to be used as the actual input to the DISE program. It is from this bulletin that the master variable file and master event common area would be created during the INIT command of DISE. However, it is again pointed out that presently the INIT command is automatically invoked and merely accesses the master variable file, which has been prepared offline.

Master Variable File

This is a random-access, unformatted file containing the variables to be used in the analysis session. The format was given in Table 2 earlier. The entries are arranged by event in this file, and pointers are maintained in the master event common area to access the data. The file length will vary, but is roughly given by #events x #stations x #variables x 22 bytes, where "#stations" implies the average number of stations per event and "#variables" implies the average number of variables per station.

Station Constant File

The station constant file contains location data on the seismic stations and descriptions of their sensors. This file is available on the VAX-11/780 and is adequate for the DISE program. The file length is presently roughly 100K bytes.

World Map File

The world map file will actually consist of many files. One large file holds the entire set of digitized world coastlines. A group of 648 separate files hold 10 degree by 10 degree segments of this world coastlines map. Similar files exist for the political boundaries. The total length of these files is approximately 130K bytes.

Data Limitations

The virtual memory feature of the VAX-11/780 obviates much of the I/O for temporary data sets required by most other existing mainframe computers. Thus, the DISE program has been coded to maintain extensive data arrays in common areas and in work areas unique to each module. However, current dimensions allow for

9 groups
9 subgroups per group
200 events per group
30 variables
30 stations
1000 events

In addition, the length of the vector holding the data arranging by subgroup as in Figure 6 is 50,000.

REFERENCES

IMSL, 1980, IMSL Libraries Reference Manual, International Mathematical and Statistical Libraries, Inc., Houston, Texas.

Ringdal, F., 1976. Maximum - likelihood estimation of seismic magnitude, Bull. Seism. Soc. Am., 66, 789-802.

Rivers, D. W., M. E. Marshall, J. A. Burnetti, R. A. Wagner, P. J. Klouda, and A. O'Donnell, An Evaluation of the VSC Discrimination Experiment (U) VSC-TR-81-12.

ILMED

8